Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Shivani Beall April 13, 2020

Development and assessment of real-time PCR assays to detect diffusely adherent *Escherichia coli*

by

Shivani Beall

Karen Levy, PhD, MPH and Jonas Winchell, PhD, MS
Advisers

Department of Biology

Karen Levy, PhD, MPH and Jonas Winchell, PhD, MS
Advisers

Katia Koelle, PhD
Committee Member

Kate O'Toole, PhD
Committee Member

2020

Development and assessment of real-time PCR assays to detect diffusely adherent *Escherichia coli*

By

Shivani Beall

Karen Levy, PhD, MPH and Jonas Winchell, PhD, MS
Advisers

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2020

Abstract

Development and assessment of real-time PCR assays to detect diffusely adherent *Escherichia coli*

By Shivani Beall

The most newly recognized and one of the least studied pathotypes of diarrheagenic *E. coli* bacteria is Diffusely Adherent *E. coli* (DAEC). While molecular assays have been developed for detection of DAEC, no quantitative real-time PCR (qPCR) assay exists. The aim of this project was to create a rapid and specific real-time PCR assay able to exclusively detect organisms that define the DAEC classification of *E. coli*. We aligned 85 DAEC whole genome sequences in order to find DNA regions exclusive to this pathotype with which to design and evaluate candidate primers and probes for a DAEC-specific qPCR assay. A thorough comparison and assessment of each experimental design was performed. By using a known DAEC strain, as well as *E. coli* strains from other pathotypes, several parameters of the assay were assessed, including: specificity, sensitivity, efficiency, and the ability to discern closely related, yet genetically distinct pathotypes. In the future, this assay may provide a useful and more convenient way to identify DAEC in clinical cases and outbreak investigations. This rapid test would also greatly enhance the understanding and knowledge base of a lesser-known pathotype that is suspected of being an important cause of diarrhea in children, especially in resource-poor areas of the world. The value of this test may be further increased if incorporated into a multi-pathogen panel of diarrheagenic microbes, particularly when used in global settings that are most affected by diarrheal illness.

Development and assessment of real-time PCR assays to detect diffusely adherent *Escherichia coli*

By

Shivani Beall

Karen Levy, PhD, MPH and Jonas Winchell, PhD, MS

Advisers

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Biology

2020

Acknowledgements

My sincere thanks to all of my committee members for taking the time to help me successfully complete this thesis. I would specifically like to thank Dr. Levy for sharing her support and expertise throughout this project and Dr. Winchell for providing me with the space, resources, and guidance necessary to complete my work.

I am additionally grateful to Dr. Kelsey Jesser and Courtney Victor from Dr. Levy's research group for sharing their insight into the subject and to Dr. Maureen Diaz, Dr. Bernard Wolff, Dr. Shatavia Morrison, and Dr. Jessica Waller from Dr. Winchell's group for providing their technical expertise and assistance.

Table of Contents

Index of Figures

Introduction

E. coli is a closely-related, albeit heterogeneous group of microorganisms consisting of several pathotypes capable of causing enteric disease [1, 2]. There are six well-defined and widely recognized diarrheagenic *E. coli* pathotypes each characterized by its own mode of pathogenesis within a host. These include enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), and diffusely adherent *E. coli* (DAEC) [1]. Of these six groups, DAEC is underrepresented in studies of diarrhea globally, although it has been confirmed as a diarrheagenic group of *E. coli* and has been specifically implicated in cases of diarrhea in young children greater than 12 months of age as well as in urinary tract infections [3-5]. More recently, DAEC has been associated with diarrheal outbreaks in urban regions of Ecuador [6].

While regional studies focused in Latin America have found DAEC to be an important factor in diarrheal disease, other large-scale global studies have not screened for this pathogen. The Global Enteric Multicenter Study (GEMS) and the Malnutrition and Enteric Disease Study (MAL-ED) are the two largest studies that have identified causal pathogenic agents of childhood diarrheal disease in sites around the globe. Both studies utilize multiplex PCR to detect multiple pathotypes of *E. coli* (along with a variety of assays for other enteric pathogens) for surveillance, however neither study has included a test for DAEC [7, 8]. As other studies are showing an increasing association of DAEC to diarrheal illness and outbreaks, the incorporation of a rapid and specific test for DAEC into large studies such as GEMS and MAL-ED would improve surveillance of enteric disease worldwide. We hope to provide such an assay for future use in a multi-pathogen panel of diarrheagenic microbes, as

improved detection of DAEC in clinical samples is key to developing a greater understanding of the prevalence of this pathotype and its potential contribution to outbreaks and sporadic disease.

Currently, a phenotypic adherence assay, which differentiates the pathotype visually by its diffuse pattern of adherence to eukaryotic cells, serves as the "gold standard" in DAEC identification [2]. The identifiable adherence pattern is due to the expression of the Afa/Dr family of adhesins in DAEC. The Afa/Dr operon family has also been used as a genetic marker in conventional PCR assays, specifically targeting the region known as AfaB-C, which encodes a conserved structural region of a large subset of Afa/Dr operons, but excludes a minority of DAEC strains [3, 4, 9]. Additionally, *E.* coli undergoes extensive horizontal gene transfer, which can occur within and between pathotypes, making genotypic characterization difficult [10, 11]. DAEC has sometimes been defined as a subclass of EAEC, as some EAEC strains produce adhesins of the Afa/Dr family, while also possessing genes for the expression of other virulence factors [1, 12]. This overlap means that existing conventional PCR assays also cannot specifically target DAEC strains, as they also detect some EAEC strains possessing the Afa/Dr operon conserved region [13].

While conventional PCR allows for qualitatively determining the presence of a gene target in a sample, a quantitative real-time PCR (qPCR) assay would greatly improve the efficiency speed, and convenience of detection as it eliminates the need for post-PCR amplification analysis such as gel electrophoresis, while also providing real-time quantitative results. Additionally, a multiplex qPCR would have the potential to test for the multiple subunits that make up the Afa/Dr operon found in DAEC. Such an assay would test for both the conserved structural regions and the variable adhesin regions that must be

expressed in order for Afa/Dr adhesins to exhibit the diffuse pattern of adherence by which DAEC is defined [3].

This project reports the identification of optimal amplicon targets for the development of a qPCR assay with the intent to validate and compare relative amplification efficiency, specificity, and sensitivity of the potential assays in the near future. Two candidate assays were assessed against extracted *E. coli* DNA of the six pathotypes as well as clinical isolates. Experiments were conducted to ensure the ideal assay is both inclusive of all DAEC yet exclusive to all other organisms, including non-DAEC *E. coli*. Eventually, this assay could be included in a larger PCR panel to analyze the prevalence of DAEC in the context of other *E. coli* pathotypes and enteric organisms.

Materials and Methods

85 whole genome sequences from DAEC isolates of the EcoZUR study were aligned along with sequences from other pathotypes in order to identify unique regions to be used as potential qPCR targets. From these regions, two candidate assays were designed based on appropriate amplicon length and oligo melting temperatures. The two assays were tested against no-template control in order to ensure no self-hybridization within the oligos, and a limit of detection analysis was also performed. The assays were then tested against DAEC isolates to test for inclusivity as well as representative isolates from other pathotypes to ensure specificity.

E. coli strains and Nucleic Acid extraction

Isolates from the Centers for Disease Control and Prevention (CDC) repository that were recovered from various sources and studies were used throughout this project.

Representative isolates of each *E. coli* pathotype were selected for initial testing for the presence of the Afa/Dr genetic region. These included 7 EAECs, 2 EIECs, 4 ETECs, 5 EPECs, 5 EHECs, and 1 DAEC, which served as the positive control.

We received additional isolates from the EcoZUR study ("*E. coli en Zonas Urbanas y Rurales*"), an epidemiological study of diarrhea in northern Ecuador [6, 14]. 65 isolates were used, including 55 DAEC, 7 ETEC, and 3 EPEC. These isolates were previously categorized by pathotype using conventional PCR assays targeting genes for virulence factors associated with each pathotype. Specifically, DAEC was identified by the presence of the Afa/Dr conserved region [6].

The EcoZUR isolates were shipped from collaborators at the Universidad de Quito (USFQ) to Emory University and transferred to the CDC on glycerol where they were stored at -80 ˚C. Once received, isolates were grown on blood agar plates and incubated overnight at 37˚C. Cultures were harvested after being sub-cultured to ensure purity and clonality. DNA was extracted using the MagNA Pure Compact Instrument and Nucleic Acid Isolation Kit (Roche Molecular Systems).

PCR and amplicon analysis

The 24 *E. coli* isolates from the CDC repository were first evaluated using a previously validated conventional PCR assay to assess the presence/absence of the Afa/Dr operon. Primers designed by Mansan-Almeida [9] that target a 750 base pair region of the AfaB-C gene were used. These primers are: 5'-CTGGGCAGCAAACTGATAACTCTC-3' and 5'-CATCAAGCTGTTTGTTCGTCCGCCG-3' (forward and reverse primers, respectively). The primers were manufactured by the Biotechnology Core Facility Branch of the CDC. The conventional PCR reaction mixture contained the following: 12.5μL of the Platinum qPCR

SuperMix-UDG (Invitrogen) 2X master mix, 0.5 µL each of 10 mM forward and reverse primers, 1.0 µL of 10 mM dNTPs, 0.5 µL of Platinum *Taq* Polymerase, and 5 µL of a 1 ng/uL stock of template with molecular-grade water making up the final volume of 25 µL per reaction well. A known DAEC strain from the CDC repository was used as the positive control, while nuclease-free water was used as the negative control. Conventional PCR was performed on the Bio-Rad PTC-0200 DNA Engine Thermal Cycler under the following cycling conditions: one cycle at 95˚C for 2 minutes followed by 45 cycles of 95˚C for 15 seconds, 60˚C for 30 seconds, and 72˚C for 60 seconds.

The PCR products were visualized through electrophoresis using the Agilent D1000 ScreenTape System on the Agilent 2200 TapeStation (Agilent Technologies).

DAEC genome sequences and analysis

Whole genome sequences of 85 DAEC isolates that were previously described as obtained from the EcoZUR study were made available by the Georgia Institute of Technology for bioinformatic analysis. These sequences were first assessed using BLAST against the GenBank database [15] to ensure the presence of the Afa/Dr operon. They were then aligned using the Mauve genome alignment program [16] and compared to representative non-DAEC genomes to find DNA regions exclusive to DAEC for potential use in assay design.

Assay design

Primers and probes were designed from newly identified DAEC-exclusive genome regions using the Realtime PCR Tool (Integrated DNA Technologies). Intermolecular and intramolecular analysis of the oligos was conducted using the Oligo Analysis Tool (Eurofins Genomics), which was also used to determine melting temperatures and predict if there was

any dimer formation potential. The primers and probes were individually assessed for specificity using BLAST tools on the NCBI website.

Assay validation

Each assay was tested with qPCR experiments on the 7500 Real-Time PCR System (Applied Biosystems), with cycling conditions as follows: one cycle at 95˚C for 5 minutes, followed by 45 cycles of 95˚C for 15 seconds and 60˚C for 60 seconds. The qPCR reaction mixture contained the following: 12.5 µL of PerfeCTa qPCR ToughMix (Quantabio), 0.5 µL each of 50 µM forward and reverse primers, 0.5 µL of 10 µM probe, 6 µL of water, and 5 µL of template, yielding a final volume of 25 µL per reaction well. 96 replicates were tested using nuclease free water as a no template control to ensure no amplification results from self-hybridization or intermolecular hybridization of primers and probes. Limit of detection was determined by testing 8 replicates of each dilution in a 10-fold dilution series of the DAEC control template ranging from 1 ng/µL to 100 ag/µL.

To ensure inclusivity and specificity of the assay, the nucleic acid extracted from the EcoZUR isolates was tested against both assays using the cycling conditions and reaction mixture described above. Thus far, 65 of the 85 DAEC isolates have been tested against our assays. Each isolate was normalized to 1 ng/ µL, yielding 5 ng of template in each reaction, and three replicates of each isolate were tested. The 55 DAEC isolates were used to test assay inclusivity, while the 10 ETEC/EPEC isolates were used to examine its specificity to the DAEC pathotype.

Results

PCR and amplicon visualization

Figure 1 shows the amplification bands following the conventional PCR experiments using the nucleic acid from representatives of the six pathotypes of *E. coli* from the CDC repositories. The expected amplicon size of 750bp could be readily seen in the DAEC positive control lane. As anticipated, some EAEC strains are known to carry the Afa/Dr conserved region and thus showed bands at the correct size (4 of the 7 analyzed EAECs). More surprising and concerning were faint to moderate bands of amplification in non-DAEC/EAEC pathotypes (1 EIEC, 3 EPECs, and 2 EHECs). The samples in question were regrown from their isolates, extracted, amplified, and visualized once more in order to verify the initial result. Upon retest, the samples displayed no bands at the 750 base pair mark. The initial preparation or PCR reaction may have been contaminated with a small amount of DAEC target, resulting in a false positive. Re-testing confirmed that no pathotype other than DAEC and some EAEC strains possessed this gene target.

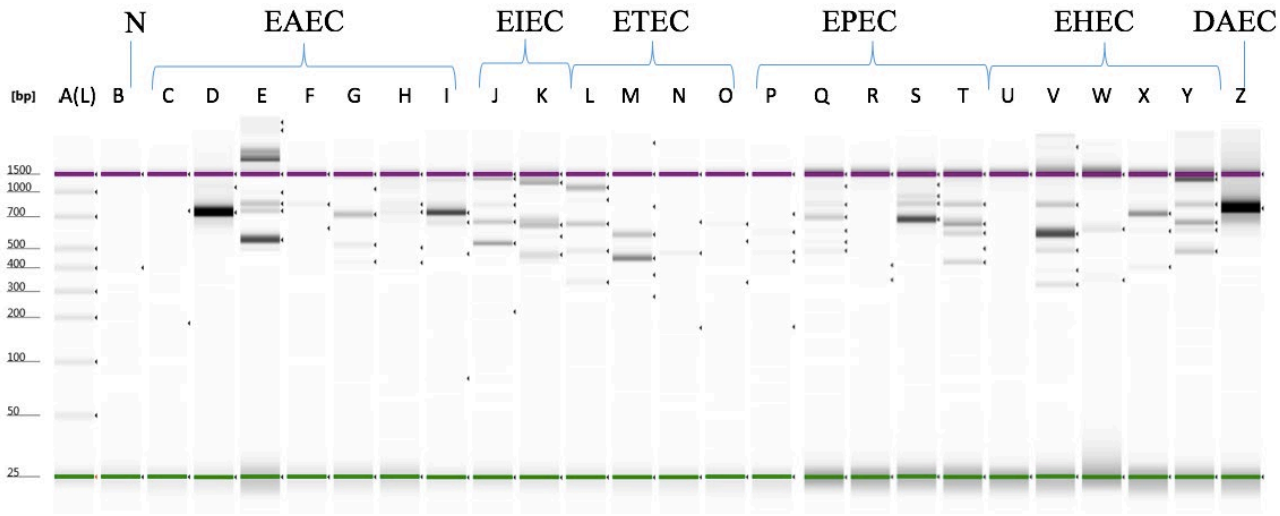

Figure 1. 2200 Tapestation (Agilent Technologies) electrophoresis of PCR amplified products. PCR products were obtained using the AfaB-C primers outlined in Materials and Methods. Lane A contains 1 kb DNA ladder and lane B is the negative control, which contains molecular-grade water in place of template DNA. The *E. coli* pathotype from which template DNA was obtained is indicated above the lanes, with DAEC serving as the positive control (lane Z).

DAEC genome sequences and analysis

From the bioinformatic analysis of 85 DAEC genomes, 5 candidate genes were identified for potential real-time PCR assay design. After a BLAST search of these genes, one was found to be present in non-DAEC *E. coli* strains, and was removed from further consideration.

Assay design

Two regions, from the PapB operon and SafD gene, both confirmed to be located within the Afa/Dr operon upon BLAST analysis, were selected for real-time assay design. Two sets of promising primers and probes were designed (Table 1), one for each target, and were evaluated *in silico* using BLAST to theoretically test whether amplification would indeed occur in only the intended DAEC strains. Additionally, appropriate melting temperatures were confirmed for each oligo. No significant evidence of self-hybridization

potential for either primer set as well as intermolecular hybridization between each forward primer and its respective probe was observed via the Oligo Analysis Tool (Eurofins Genomics). Figure 2 illustrates the location of the primers and probes relative to each other in their respective genes.

Table 1. Real-time PCR assays designed for identification of DAEC

Target Gene	Oligo Designation	Nucleotide Sequence (5' to 3')	Amplicon Size (bp)	Melting Temp. (˚C)
PapB	ECPapB-F ECPapB-R ECPapB-P	GATATCGCCGGTACACAGTG GACTGACCCGCTGAAACC CGCTGAGGGATTATCTGGTAATGGGA	146	62.45 60.16 66.47
SafD	ECSafD-F ECSafD-R ECSafD-P	CATGAGCTTCGTGTCAGGAC TTCACCAGGAGCAATGTCC TGAAGGGAGAAGGCGGGAAGG	144	62.45 62.18 66.17

A)

GATATCGCCGGTACACAGTGAGAAGGTT
ATTAATGCGCTGAGGGATTATCTGGTAA
TGGGATATAACCGCATGGAGGCCTGCGG
GCGTCATAGTGTGTCGCCGGGATATTTT
TCTGGTGCACTGAAGCGGTTTCAGCGGG
TCAGTC

B)

CATGAGCTTCGTGTCAGGACAGGAGGAG
ATGGCTGGTCGCCGGTGAAGGGAGAAGG
CGGGAAGGGGGTGTCGAGGCCCGGTCAG
GAGGAGCAGTTTTTTTTTGATGTGATGG
CGGACGGAAATCAGGACATTGCTCCTGG
TGAA

Figure 2. Locations of primers and probes within the PapB (A) and SafD (B) amplicons. DNA is shown in the 5' → 3' direction. Sequences highlighted in green indicate primers, while sequences highlighted in purple indicate the probe.

Analytical validation

No amplification was observed with either assay when tested with nuclease free water as a no-template control, indicating no self-hybridization or intermolecular hybridization of the primers occurred during the reaction. Additionally, the limit of detection was determined to be 10 fg/µL for each assay as defined by having at least half the replicates displaying amplification (Table 2). Since 5 µL of template was used in each reaction, the limit

of detection for both the PapB and SafD assays is 50 fg. At concentrations below 10 fg/µL, no amplification was observed in any replicates.

Table 2. Summary of limit of detection for PapB and SafD assays

	PapB			SafD		
	Avg. Ct	**Std. Dev.**	**Reps Pos**	**Avg. Ct**	**Std. Dev.**	**Reps Pos**
1 ng/µL	22.39	0.25	8/8	21.40	0.16	8/8
100 pg/µL	25.47	0.17	8/8	24.49	0.25	8/8
10 pg/µL	29.25	0.42	8/8	27.17	0.18	8/8
1 pg/µL	32.91	0.31	8/8	30.34	0.15	8/8
100 fg/µL	36.49	0.39	8/8	33.68	0.58	8/8
10 fg/µL	39.65	1.24	6/8	37.22	0.95	7/8
	LOD	50 fg		**LOD**	50 fg	

Specificity

Of the 55 EcoZUR DAEC isolates tested, 47 exhibited strong amplifications in all three replicates against both assays. The 8 other isolates showed no amplification in either of the assays. Since all isolates were screened with conventional PCR for the presence of a conserved region of the Afa/Dr operon, further genomic analysis was conducted to determine if these 8 isolates possess all the genomic markers within the Afa/Dr operon that would classify them as DAEC. Upon re-analysis, 4 of the 8 isolates did not display both the conserved structural regions and variable adhesin regions necessary for the diffuse adherence pattern associated with DAEC. Therefore, only 4 of 55 isolates considered to be DAEC went undetected by our assays. For the majority of DAEC isolates, the consistency of detection sensitivity for each newly designed marker was demonstrated.

Nucleic acid template from the 10 non-DAEC isolates (7 ETECs and 3 EPECs) exhibited no amplification in either of the assays, indicating the selected targets are not

present in ETEC and EPEC pathotypes as expected. Isolates from the other four *E. coli* pathotypes would also need to be screened to ensure specificity.

Discussion

The goal of project was to develop a specific real-time PCR assay to detect DAEC, a lesser-known, but potentially significant diarrheagenic pathotype of *E. coli*. While ways to discern between pathotypes exist, through conventional PCR and phenotypic adherence assays, quantitative real-time PCR could quickly, accurately, and more efficiently detect the presence of DAEC. Such an assay could prove useful, especially if used in conjunction with assays for other *E. coli* types and enteric pathogens to assess the prevalence of DAEC.

Assay Development

The first step of the project was to assess if the Afa/Dr operon conserved region known as AfaB-C could be considered a defining genotypic characteristic of DAEC. As DAEC is a subclass of EAEC, some EAEC strains do possess genes encoding Afa/Dr adhesins in addition to genes for other virulence factors that characterize EAEC. Therefore, it is entirely possible that some (or all) EAEC samples would possess this target and appear positive upon PCR testing. After screening all test specimens, an analysis of the PCR products (after the retest of EIEC, EHEC, and EPEC samples that appeared to display bands potentially from the PCR target) showed the 750 base pair region from AfaB-C appeared to be present in the DAEC sample as well as in a few of the EAECs.

After the two regions (PapB and SafD) were selected for assay design from whole genome sequence alignment, BLAST searches indicated they are both located within the Afa/Dr operon family. Although the initial bioinformatic analysis indicated these regions

were not present in the representative EAEC assemblies, laboratory testing of the assays with qPCR using EAEC isolates as templates may result in fluorescence signals if these regions are present in EAECs that possess Afa/Dr adhesin genes.

Assay Performance

Upon validation of the assays, both exhibited equal performance in response to negative controls (no self-hybridization) as well as the DAEC positive control. With each assay having a limit of detection of 50 fg, they were used against DAEC and non-DAEC isolates to test for inclusivity and specificity. Again, both performed equally, detecting 47 of the 55 DAECs and none of the isolates from other pathotypes. Further genomic analysis of the DAEC isolates that appeared negative is required, along with future experiments against isolates in the EAEC, EIEC, and EHEC pathotype categories. Testing thus far, however, demonstrates that the two assays show consistent and equal performance with *E. coli* isolates.

The definition and agreed upon genetic markers for DAEC detection are continually being updated. Previous literature has shown that screening for conserved regions of the Afa/Dr operon family may not be sufficient to detect all DAECs with no overlap into other pathotypes, such as EAEC [3, 13]. By conducting whole genome alignment, we screened for regions of the genome that were representative and exclusive to DAEC irrespective of their location within the Afa/Dr operon family. This method highlighted genome regions that are highly conserved across a representative sample of DAECs. The qPCR assays may therefore be more inclusive than the currently available conventional PCR that detects only a subset of the Afa/Dr operon family [3, 9]. Nevertheless, further analysis may reveal that a more intricate multiplex assay is necessary in order to detect all DAEC strains while excluding non-DAECs that may contain conserved Afa/Dr operon regions.

Limitations and Future Work

One notable limitation of this project is the circular nature of assay design and validation. The isolates obtained from the EcoZUR study were initially classified as DAEC by screening for the Afa/Dr operon via conventional PCR. Then, whole genome sequences from these Afa/Dr-positive isolates were used to identify unique genomic regions from which qPCR assays were designed. These assays were then tested against the isolates from which these whole genome sequences came. The assays have therefore not been tested against isolates whose genomes were not used in their design and that were not previously screened for the Afa/Dr operon. In the future, we may screen against DAEC isolates identified by the phenotypic adherence assay in order to ensure that the qPCR assays are not solely effective against the isolates from which their DNA targets were identified and derived.

Additionally, specificity to the DAEC pathotype has not been fully assessed because only representative isolates from the ETEC and EPEC pathotypes have been screened. We plan to conduct additional testing against the non-DAEC isolates in the CDC repositories in order to ensure assay specificity. This testing would be particularly important to screen the EAEC isolates that are known to possess conserved regions of the Afa/Dr operon.

Beyond testing the assays against a broader range of non-DAEC isolates, the ultimate goal is to test these newly designed assays on specimens from clinical cases. There is currently an effort to obtain clinical specimens from the same partner in Ecuador in order to further assess the utility of the candidate assays. Ideally, this assay, combined with others that detect different *E. coli* strains and possibly other pathogenic enteric bacteria, would be used in a primary testing facility to provide a rapid and accurate diagnosis for patients who show clinical symptoms consistent with infection with these organisms. From a research

and epidemiology standpoint, this DAEC assay would be a useful tool for surveillance studies to understand the role that this pathotype may play in disease in certain populations and provide valuable data on the prevalence and potential significance of this group of organisms.

References

1. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
2. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli.* Clin Microbiol Rev, 1998. **11**(1): p. 142-201.
3. Servin, A.L., *Pathogenesis of human diffusely adhering Escherichia coli expressing Afa/Dr adhesins (Afa/Dr DAEC): current insights and future challenges.* Clin Microbiol Rev, 2014. **27**(4): p. 823-69.
4. Le Bouguenec, C. and A.L. Servin, *Diffusely adherent Escherichia coli strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens.* FEMS Microbiol Lett, 2006. **256**(2): p. 185-94.
5. Scaletsky, I.C.A., et al., *Diffusely adherent Escherichia coli as a cause of acute diarrhea in young children in northeast Brazil: a case-control study.* Journal of Clinical Microbiology, 2002. **40**(2): p. 645-648.
6. Pena-Gonzalez, A., et al., *Metagenomic Signatures of Gut Infections Caused by Different Escherichia coli Pathotypes.* Appl Environ Microbiol, 2019. **85**(24).
7. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* The Lancet, 2013. **382**(9888): p. 209-222.
8. Platts-Mills, J.A., et al., *Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED).* The Lancet Global Health, 2015. **3**(9): p. e564-e575.
9. Mansan-Almeida, R., A.L. Pereira, and L.G. Giugliano, *Diffusely adherent Escherichia coli strains isolated from children and adults constitute two different populations.* BMC Microbiology, 2013. **13**(22).
10. Leimbach, A., J. Hacker, and U. Dobrindt, *E. coli as an all-rounder: the thin line between commensalism and pathogenicity.* Curr Top Microbiol Immunol, 2013. **358**: p. 3-32.
11. Lopes, L.M., et al., *Heterogeneity among strains of diffusely adherent Escherichia coli isolated in Brazil.* J Clin Microbiol, 2005. **43**(4): p. 1968-72.
12. Palaniappan, R.U., et al., *Differentiation of Escherichia coli pathotypes by oligonucleotide spotted array.* J Clin Microbiol, 2006. **44**(4): p. 1495-501.
13. Snelling, A.M., et al., *The commonly-used DNA probe for diffusely-adherent Escherichia coli cross-reacts with a subset of enteroaggregative E. coli.* BMC Microbiol, 2009. **9**: p. 269.
14. Smith, S.M., et al., *Locals get travellers' diarrhoea too: risk factors for diarrhoeal illness and pathogenic Escherichia coli infection across an urban-rural gradient in Ecuador.* Trop Med Int Health, 2019. **24**(2): p. 205-219.
15. Altschul, S.F., et al., *Basic Local Alignment Search Tool.* J. Mol. Biol., 1990. **215**(3): p. 403-410.
16. Darling, A.C., et al., *Mauve: multiple alignment of conserved genomic sequence with rearrangements.* Genome Res, 2004. **14**(7): p. 1394-403.